N66-15236

The Origin of the Moon

Paul D. Lowman, Jr.
Goddard Space Flight Center
Greenbelt, Maryland

Introduction

The origin of the moon is one of the most refractory problems of cosmogony, despite the fact that we know more about the moon than about any other member of the solar system (excluding the earth). There are several reasons for this paradox. A major one is our ignorance of the moon's chemical and petrologic composition; we have actually learned more of stellar compositions, by spectroscopy, than of the moon's composition. In addition, we know very little of the moon's internal structure.

A more general reason for the difficulty of the problem is the fact that the moon is part of the solar system's only binary planet system. Furthermore, the earth-moon system is quite different from the other planet-satellite groups in the relative sizes of its members. Whereas most planets have masses a thousand times that of their combined satellites, the earth's mass is only eighty-one times that of the moon. The disparity in volumes is of course even more striking: the solar system's biggest satellite, Triton, has a diameter only a thirtieth that of Neptune, but the moon's diameter is a quarter that of the earth. (It should be pointed out, however, that the low densities of the giant planets indicate that their rocky cores, if any, are relatively small; the real disparity may be less than indicated.)

FACILITY FORM 602

N66-15236 (ACCESSION NUMBER) — 1 (THRU)

20 (PAGES) — 1 (CODE)

TMX 56134 (NASA CR OR TMX OR AD NUMBER) — 30 (CATEGORY)

[redacted]

Another fact which emphasizes the unique nature of the earth-moon system is the low density (3.34 g/cm^3) of the moon, compared to those of the other planets and the chondritic meteorites (all in g/cm^3): Earth, 5.5; Mars, 4.2; Mercury, 5.0; Venus, 4.9; chondritic meteorites, 3.4-3.8.

The moon, then, is doubly unique, both by itself and as a member of a unique system. This fact has given rise to a markedly divergent group of theories as to its origin. Before discussing these theories, however it is helpful to review certain aspects of the broader problem of the origin of the planets, and in particular to summarize a few widely-accepted generalizations in this area.

First, it appears that the formation of planetary systems is a fairly common occurrence, although no others have yet been detected directly. Evidence for this belief includes the systematic deficiency in angular momentum of the dwarf stars (of which the sun is one), compared to the more massive stars, which suggests strongly that the missing momentum is tied up in unseen planetary systems (Struve, 1950). In addition, an unseen sub-stellar companion of 61 Cygni has been detected by its gravitational effects (Strand, 1957).

A second generalization is that the formation of the solar system was essentially monistic, i.e., it required no intervention of external bodies or forces as did, for example, the Chamberlain-Moulton tidal theory.

Finally, most modern theories of the origin of the planets hold it to have been, at least initially, a process of low-temperature

[redacted]

accretion of solid material, rather than condensation of hot solar material. Details of the accretion process, however, are extremely difficult to reconstruct.

Turning now to the moon, we see that current theories of its origin utilize three major mechanisms: independent formation near the earth, fission of the earth, and capture of one or more bodies by the earth.

Independent Origin

In this concept, variations of which have been proposed by Kuiper (1951), Opik (1961), and Ruskol (1962), the moon is supposed to have been formed by accretion relatively close to the earth while the earth itself was still growing. There are significant differences in detail between the mechanisms proposed. Kuiper suggests that the earth and moon grew from a double proto-planet, whereas Ruskol, to explain the disparity in masses, suggests capture of many small bodies from the protoplanetary cloud in the earth's orbit, which accumulated to form the moon. There is general agreement that the pre-mare lunar craters, and perhaps the maria themselves, represent the last stages of the accumulation process.

The binary system concept has several attractive features. Ruskol, for example, cites the "regular character" of the moon's orbit, contrasted with that of retrograde satellites such as Triton and Phoebe, as evidence of the moon's formation near the earth. Furthermore, MacDonald (1964) finds the dynamical history of the earth-moon system to be consistent with such an origin, except for the time required (a difficulty discussed below). Finally, if we accept Kuiper's proposal that the earth-moon system formed in the same way as do binary stars, there is no problem in finding analogues to it.

The major weakness of the independent origin theories is their failure to explain convincingly the difference in mean density between the earth and the moon. This density difference is generally believed to represent a difference in composition, specifically in the proportions of metallic iron and silicates, and it is difficult to see why objects accumulating in the same part of the primitive solar nebula should have such different compositions. Kuiper (1952) suggests that if the earth began to form somewhat earlier than the moon, evaporation of silicates from the earth and subsequent removal of the oxides by radiation pressure might have enriched the earth in iron compared to the moon. Alternatively, selective accretion of iron might be responsible. Neither of these explanations has been generally accepted, however. A third possibility, endorsed by Ruskol, is that the density difference does not reflect a compositional difference, but is instead the result of pressure-(and hence size)- dependent phase changes. This explanation was proposed to explain formation of the earth's core by Ramsey (1948), but has been almost completely invalidated by more recent data on the density of Mercury and by much independent geophysical evidence (Wildt, 1963; Birch, 1961); the nearly-unanimous concensus now is that the core is essentially iron.

Another weakness in the independent origin concept has been pointed out by MacDonald (1964), who finds that it would take more time than seems to be available for a binary system to evolve to the present configuration of the earth-moon system unless the present tidal interactions are abnormally strong. Independent evidence of the rate of change in length of day indicates that such is not the case, however; the time-scale difficulty is thus not met.

Fission of the Earth

As a result of his extensive studies of the tidal interactions of the earth, moon, and sun, G. H. Darwin (1879) concluded that the earth and the moon might have once formed one body, rotating in about five hours. The period of free oscillation of a homogeneous fluid earth had been calculated by Thomson to be 1 hour and 34 minutes, and would probably be slightly greater if the earth possessed a core, so that about two oscillations could be completed in one rotation. Darwin suggested that the solar tides, reinforced by the coinciding free oscillations and the rapid rotation of the primitive earth, might have led to fission into one large and one or more smaller bodies, the latter eventually forming the moon.

Darwin's "wild speculation," to use his own term (1901), has had a long and interesting history since its birth. More detailed study of the problem by Moulton, Brown, Nölke, Lyttleton, and Jeffreys, as summarized by Jeffreys (1959), brought about the nearly complete rejection of the fission mechanism, but modified versions have more recently been proposed by Wise (1963), Cameron (1963), and O'Keefe (1963).

Wise and Cameron advocate fission resulting from the increased rotational rate attendant on segregation of the earth's core. We shall review briefly the various arguments.

The strongest support for Darwin's hypothesis comes from the study of lunar motions, which shows that the moon must at one time have been closer to the earth, with a correspondingly shorter month and day. Extrapolations of this sort cannot, however, be extended rigorously to the time when the earth and moon were in contact, as Darwin stressed. The coincidence of the periods of free oscillation and solar tides, if the earth rotated in four or five hours, furnishes *a priori* support for the theory, and the combined angular momentum of the earth and moon would correspond to a single body rotation period of about four hours (Jeffreys, 1959).

Certain features of the earth may represent physical evidence of the moon's formation by fission. The most obvious of these is the Pacific basin, which was suggested by Pickering to be the scar left by removal of the primitive crust during fission; Wise suggests that this material may now constitute the far side of the moon. Wise points out, however, that modern theories of continental genesis tend to remove the need for such an explanation, since the continents are now generally thought to have grown by accretion over geologic time. Much less obvious evidence of the moon's birth may be the non-hydrostatic shape of the earth (the "pear-shaped" component) discovered by O'Keefe and Eckels (1958) from analysis of the motion of Vanguard I. O'Keefe (1963) points out that this shape can best be explained by supposing that small but significant stresses can be supported indefinitely by the mantle, i.e., that it has a finite strength, and that these stresses may have been the result of fission of the earth.

Cameron suggests that the composition of the earth's atmosphere is evidence for formation of the moon by fission. The terrestrial atmosphere cannot represent a residue of the primordial solar nebula, as can that of Venus; the difference between the atmospheres of such similar planets may be, according to Cameron, the result of loss of the primordial terrestrial atmosphere during fission.

Some of the strongest independent support for the fission theory is the near coincidence of the density, and by inference, the composition, of the moon and the earth's upper mantle (3.34 g/cm^3 for the moon, 3.3 to 3.9 g/cm^3 for the mantle above 500 km. (Clark and Ringwood, 1964)), which are explainable if fission followed core formation, as suggested by Wise and Cameron. Further independent evidence may come from tektites, although their lunar origin is still disputed. Tektites are chemically similar to the average continental crust of the earth, not only in bulk composition, but also in the relatively low Ni/Fe ratio. The latter can be explained by core formation in the earth, but it is clear that the moon possesses no large core. Tektite chemistry, then, tends to support the fission theory, if they are of lunar origin, because both the moon and the earth's crust would be derived from the mantle under this theory.

Arguments against the fission theory have been developed by several scientists. Jeffreys (1930) found that the earth's core would cause considerable internal friction, leading to such rapid dissipation of energy in the oscillating earth that the bulge so produced could not be greater than about 1/23 of the radius - clearly far short of the hour-glass-shaped Poincare figure suggested by Darwin (1901).

Another difficulty is the high rotation rate required for fission. The angular momentum density of the planets appears to be proportional to their mass (MacDonald, 1964), with the exception of Mercury and Venus, which have been slowed by tidal friction. The rotation rate necessary to produce fission in the earth would correspond to the angular momentum density of a planet with ten times the earth's mass. To postulate fission then requires the ad hoc assumption of an unusually high rotation rate, an unsatisfactory situation even if one postulates, as does Wise, loss of angular momentum from the earth-moon system by magnetic drag.

MacDonald suggests another weakness in the fission theory on the basis of his study of the dynamical history of the earth-moon system. The fragments ejected from the earth would be in the equatorial plane, and hence tidal interactions would not tend to change the inclination of the moon's orbit. However, MacDonald demonstrates that the inclination (and hence the obliquity of the earth's equator) has changed gradually by tidal interaction over geologic time; the fission theory therefore cannot provide the necessary initial conditions inferred for the earth-moon system.

Capture by the Earth

The apparent difference in bulk composition between the earth and the moon figures importantly in all theories of the moon's origin. As stated previously, it constitutes a major difficulty for the independent accretion mechanism, at least if the accretion was near the earth. Urey (1962) accordingly suggested that the moon was captured by the earth; variations of the capture mechanism have also been suggested by Gerstenkorn (1955), Alfvén (1963), and MacDonald (1964). Because of their complexity, these theories will be reviewed separately.

Urey's theory for the origin of the moon forms part of his overall concept of the evolution of the planets, a full discussion of which would be beyond the scope of this article. Briefly, Urey proposes that the first large solid bodies ("primary bodies") to form in the solar system were approximately the size of the moon and had the composition of achondritic and iron meteorites. These bodies collided, due to gravitational perturbations, over a space of about 200,000,000 years, the fragments re-accumulating to form the planets. The moon is considered to be one of these primary bodies which escaped destruction and was captured by the earth.

Urey's capture mechanism has the advantages of explaining certain properties of meteorites and the differing compositions of the moon and the terrestrial planets. In addition, Urey points out that the capture of the moon does not seem improbable if it is viewed as the survivor of many such captures.

Several weaknesses in this mechanism apply in some degree to other capture theories. MacDonald (1964) finds that the time available is insufficient for evolution of the present earth-moon system if it started by capture, a difficulty shared with the independent origin theory. An additional problem is the earth's thermal history (MacDonald, 1959); the present heat flow is consistent, within a factor of two, with a chondritic composition, and there seems no way to dispose of the great amount of heat which would be generated in the body of the earth by capture-induced body tides. Finally, Urey's theory is partially based on certain controversial assumptions as to the origin of meteorites. For example, the necessity for lunar-size objects is partly nullified by the discovery by Lipschutz and Anders (1961) that at least some of the diamonds in meteorites have been formed by impact, rather than by static pressure. Also, the Cumberland Falls achondrite contains fragments of a chondritic meteorite, or precisely the reverse of the age relationship predicted by Urey's theory.

Gerstenkorn (1955) proposed a theory, later adopted and modified by Alfvén (1963), in which the moon was originally another planet captured by chance by the earth. The event changed the moon's orbit to a retrograde one inclined 149° with respect to the earth's equator. This orbit changed as the moon came closer, so that it began to circle the earth in the same direction as the earth's rotation. At the minimum distance, 2.89 earth radii, the inclination was 45.7° with respect to the earth's equator. Alfvén points out that this minimum distance coincides with the Roche limit, and that the moon may therefore have broken up, part of the debris falling on the earth. This material formed the nuclei for the continents, or even all of the earth's crust above the Mohorovicic discontinuity.

Inasmuch as it is derived from studies of tidal interaction, the Gerstenkorn-Alfvén theory accounts, at least in principle, for the present dynamical state of the earth-moon system. Furthermore, it tends to meet the composition problem by relegating the moon's origin to a different region of the solar system; Alfvén's earlier work (1954) treated the moon as an independent planet whose composition was the result of the magnetic sorting of ionized gas.

Other aspects of the Gerstenkorn-Alfvén theory, however, appear to have significant weaknesses. In addition to those common to any single-body capture mechanism, discussed previously, the suggestion that the earth's continents represent former lunar material accumulated after the Roche limit break-up is open to criticism. The continents have a mean density of 2.7-3.0 g/cm^3, if we include all the material above the Mohorovicic discontinuity. However, the moon's density is essentially equal to that of the earth's upper mantle. The density difference between the crust and mantle is of great significance, representing, regardless of the nature of the Mohorovicic discontinuity, extremely different rock types. The continents could not be formed, therefore, by addition of lunar material to the earth, unless we suppose the moon to have been differentiated before capture, in which case the earth itself could have been also differentiated by purely internal processes.

MacDonald (1964) has suggested a variation of the capture theory in an attempt to conform to the limits set by his studies of dynamical history of the earth-moon system. He suggests that the moon originated by the mutual collision and accretion of several smaller moons which had previously been captured or formed near the earth. This process would be dominated by the largest of these moons, and would avoid the time scale problem of single-body capture. He points out, however, that multi-body capture is still a largely uninvestigated subject.

Summary

Unlike theories of the origin of the solar system, theories of the moon's origin have been diverging radically in recent years. Even the oldest concepts, such as Darwin's tidal fission theory, are retained as working hypotheses because of apparent crucial weaknesses in their successors, and completely new theories are still being developed. It is apparent, then, that the moon's origin remains one of the most difficult problems of cosmogony. We shall discuss briefly a few of the most promising lines of investigation which may ultimately lead to its solution.

Theoretical studies of the origin of the solar system rank high among these, since it is clear that the moon must be viewed in context. The rapid evolution of this field promises to accelerate, and should eventually illuminate the problem of the moon's origin in ways now unforeseen.

Current studies of the dynamical history of the earth-moon system, such as those of MacDonald and Gerstenkorn, are of great potential value, being founded to a considerable degree on objectively verifiable quantities such as changes in the length of day. Advances can therefore be expected both from the continuing growth of knowledge and from improved use of this knowledge.

An obvious approach to the problem of the moon's origin is the study of lunar geology. Greatly advanced in recent years because of the realization of manned space flight, the achievement of manned lunar exploration will begin the direct interpretation of the moon's geologic record. In the course of such exploration, gaps in our knowledge of the composition of the solar system will be narrowed, with immediate application to the question of the moon's origin.

A closely related approach is one suggested long ago by Darwin: namely, the study of the geological evidence on earth bearing on changes in the moon's distance. The immense increase in our knowledge of Precambrian geology in particular should throw light on this problem, because the third-power dependence of tide-raising force on distance means that the geologic effects of a much closer moon should be substantial. Further geophysical discoveries, such as that of the earth's non-equilibrium figure by Vanguard I observations (O'Keefe and Eckels, 1958), may also be applied to studies of the moon's origin.

References

Alfvén, H., "On the Origin of the Solar System," Oxford University Press, 1954.

------, The early history of the moon and the earth, Icarus, 1, 357-363, 1963.

Birch, F., Composition of the earth's mantle, Geophysical Journal, 5, 295-319, 1961.

Cameron, A.G.W., The origin of the atmospheres of Venus and the earth, Icarus, 2, 249-257, 1963.

Clark, S. P., Jr., and A. E. Ringwood, Density distribution and constitution of the mantle, Reviews of Geophysics, 2, 1, 35-88, February, 1964.

Darwin, G. H., On the precession of a viscous spheroid, and on the remote history of the earth, Philosophical Transactions of the Royal Society, Part II, V. 170, pp. 447-530, 1879.

-----, "The Tides," 2nd ed., John Murray, London, 1901.

Gerstenkorn, H., Über Gezeitenreiburg beim Zweikörper problem, Zeitschrift fur Astrophysik, 36, 245-274, 1955.

Jeffreys, H., The resonance theory of the origin of the moon (second paper), Monthly Notices Royal Astron. Soc., XCI, I, 169-173, 1930.

------, "The Earth," 4th ed. Cambridge University Press, New York, 1959.

Kuiper, G. P., Ch. 8 in "Astrophysics," ed. by J. A. Hynek, McGraw-Hill, New York, 1951.

------, in "The Atmospheres of the Earth and Planets," 2nd ed., ed. by G. P. Kuiper, U. of Chicago Press, Chicago, 1952.

Lipschutz, M. E., and E. Anders, On the mechanism of diamond formation, Science, 134, 2095-2099, 1961.

MacDonald, G.J.F., Calculations on the thermal history of the earth, J. Geophys. Res., 64, 1967-2000, 1959.

--------, Earth and moon: past and future, Science, 145, 3635, 881-890, 28 Aug. 1964.

O'Keefe, J. A., Two avenues from astronomy to geology, in "The Earth Sciences," ed. by T. W. Donelly, Rice University, Houston, 1963.

O'Keefe, J. A., and Eckels, A., Harvard College Observatory Announcement Card 1420, 1958.

Opik, E., Tidal deformations and the origin of the moon, Astron. Jour., 66, 2, 60-67, 1961.

Ramsey, W. H., On the constitution of the terrestrial planets, Monthly Notices Royal Astron. Soc., 108, 406-413, 1948.

Ruskol, E. L., Ch. 17, in "The Moon," ed. by Z. Kopal and Z. K. Mikhailov, Academic Press, New York, 1962.

Strand, K. Aa., The orbital motion of 61 Cygni, Astron. Jour., 62, 35, 1957.

Struve, O., "Stellar Evolution," Princeton U. Press, Princeton, N.J., 1950.

Urey, H. C., Ch. 16, in "The Moon," ed. by Z. Kopal and Z. K. Mikhailov, Academic Press, New York, 1962.

Wildt, R., Ch. 5 in "Planets and Satellites," ed. by G. P. Kuiper and B. M. Middlehurst, U. of Chicago Press, Chicago, 1961.

Wise, D. U., An origin of the moon by rotational fission during formation of the earths core, J. Geophys. Res., 68, 1547-1554, 1963.